REMARKS

In the patent application, claims 1, 3-40 are pending. In the office action, all pending claims are rejected.

Applicant has amended claims 19 and 31 for formal matters, and added new claims 41-48. The support for claims 42 and 44 – 48 can be found in Figure 3, and on p.12, lines 29 – 32. The support for claims 41 and 43 can be found in Figures 5 and 6, and on p.14, line 26 to p.15, line 6. As disclosed, the speech is synthesized for each segment in the decoder based on the number of frame k and the parameter x. The parameter x is selected according to given criteria. Here no waveform matching is carried out because the time synchrony between coder input and output is lost.

No new matter has been introduced.

At section 4, claims 1, 3-14, 19-40 are rejected under 35 U.S.C. 102(b) as being anticipated by *Gersho et al.* (U.S. Patent No. 6,311,154, hereafter referred to as *Gersho*). The Examiner states that *Gersho* discloses an encoding method as claimed. The Examiner considers a CELP-type encoder, such as an AbS coder, as a parametric coder which can be used for segmenting, coding and decoding audio signals.

In particular, the Examiner states that *Gersho* discloses an encoding method wherein samples of a speech signal is partitioned in the frames based on the classes of the speech signal. The Examiner points to col. 4, lines 25-27 to support the Examiner's exertion.

It is respectfully submitted that at col.4, lines 23 –34, *Gersho* discloses a method for coding a speech signal with the following steps:

a. partitioning samples of a speech signal in frames;
b. deriving a residual signal for each frame;
c. classifying the speech signal in each frame into a plurality of classes;
d. identifying the location of at least one window in the frame by examining the residual signal for the frame;
e. encoding the excitation for the frame using one of a plurality of excitation coding techniques selected according to the class of frames; and

f. confining all or substantially all of non-zero excitation amplitudes to lie within the windows.

From the above-description, it is clear that *Gersho* already determines where to set the boundary of each segment when partitioning the speech samples without knowing the audio characteristics of the speech signal in the segments. *Gersho* partitions the speech sample into frames. After partitioning, *Gersho* classifies the speech signal in each frame into classes. In other words, in *Gersho*, the samples of speech signal are first partitioned into frames and each frame is then classified into one of a plurality of classes. Before classifying, it is impossible to partition the speech signal based on the classes. Thus, *Gersho* does not partition the speech signal in frames based on the classes as stated by the Examiner. The partitioning step in *Gersho* is carried out independently of the audio characteristics of the speech signal.

In contrast, according to the present invention, the speech signals are partitioned into segments based on the audio characteristics in the speech signals. How the speech signals are partitioned depends on the audio characteristics of the speech signal. Because the audio characteristics of the audio signal varies from sample to sample, the boundary of the segments cannot be pre-determined. As a result, a segment can be long or short; it can be 10 frames or 28 frames (see Figure 3). In *Gersho*, the length of each partitioned "segment" is the same.

Thus, *Gersho* does not disclose or even suggest segmenting the audio signal into a plurality of segments based on the audio characteristics of the audio signal. For this reason alone, the invention as claims in claims 1, 19, 22, 26, 27, 31 and 32 is clearly distinguishable over the cited *Gersho* reference.

Furthermore, Gersho is irrelevant to the present invention, because the coding method in *Gersho* is completely different from the coding method of the present invention as claimed. The claimed invention is concerned with a parametric-type encoding method, whereas *Gersho* is concerned with a CELP-type encoding method.

In the parametric-type encoding method, a parametric speech production model is used to obtain a set of parameters from the audio signal so as to produce a further audio signal in the decoder based on the parameters. The parametric-type encoding and decoding method, as disclosed in the specification, does not rely on the waveform of the speech signal segments. In

fact, due to the loss of the synchrony between the coder input and output signal, waveform matching is not carried out.

A CELP coder is an example of an Analysis-by-Synthesis (AbS) coder (see col.1, line 54 to col.2, line 1). As known in the art, a CELP coder performs waveform matching on the coder output using code excitation candidates and selecting the one minimizing given error criteria. As disclosed in *Gersho*, the CELP coder relies on the residual and excitation models. *Gersho*'s coder is not a parametric coder as disclosed in the present invention.

For the above reasons, claims 1, 19, 22, 26, 27, 31 and 32 are clearly distinguishable over the cited *Gersho* reference.

As for claim 4, the Examiner states that *Gersho* discloses the characteristics including energy characteristics in the segments (col. 4, lines 65-67).

It is respectfully submitted that at col.4, lines 65-67, *Gersho* discloses that the presence of energy peaks in the smoothed contour of the residual signal is used to identify the location of window in step d. This step is carried out after the samples of speech signal have been partitioned in step a and the frame has been classified in step c. The energy peaks are used to identify the location of the windows for use in step f. Thus, energy characteristics are not used as a class basis for partitioning. *Gersho* does not disclose or even suggest partitioning the speech signals into segments based on the energy characteristics of the speech signal. When the energy peak is determined, the speech signal is already partitioned into frames.

As for claim 5, the Examiner states that *Gersho* discloses the characteristics including pitch characteristics in the segments (col. 4, lines 59-61). It is respectfully submitted that at col.4, lines 59-61, *Gersho* discloses that the frames are further partitioned into sub-frames, and at least one window within each sub-frame is positioned at a location that is a function of a pitch of the frame. Thus, the function of a pitch is used to identify the location of window in step d. This step is carried out after the samples of speech signal have been partitioned in step a and the frame has been classified in step c. As with the energy peaks, the function of pitch is used to identify the location of the windows so that all of the non-zero excitation is confined within the windows. Thus, the function of a pitch is not used as a class basis for partitioning. *Gersho* does not disclose or even suggest partitioning the speech signals into segments based on the pitch

characteristics of the speech signal. When the location of the window is located, the speech signal is already partitioned into frames.

As for claim 8, the Examiner states that *Gersho* discloses that a plurality of voicing values are assigned to the voicing characteristics and the segmenting is carried out based on the assigned voicing values (col. 4, lines 52-53). It is respectfully submitted that, at col. 4, lines 52-53, *Gersho* discloses that the step of classifying uses a first classifying to classify a frame as being one of an unvoiced frame or a not-unvoiced frame. First, *Gersho* does not suggest assigning voicing values to the voice characteristics. Second, *Gersho* carries out the step classifying after partitioning. Thus, *Gersho* does not disclose or even suggest that the segmenting is carried out based on the assigned voicing values.

As for claims 9 to 11, they are dependent from claim 8 and recite features not recited in claim 8. For reasons regarding claim 8 above, it is respectfully submitted that *Gersho* does not disclose or even suggest the features in claims 9 to 11.

As for claim 12, the Examiner states that *Gersho* discloses a further step of selecting a quantization mode such that the segmenting step is carried out based on the selected quantization mode (col. 3, lines 45-59; Figure 5 and col. 11, lines 4-16; col. 4, lines 36-37; col. 15, lines 35-36 and col. 9, lens 63-65). It is respectfully submitted that *Gersho* discloses a method of speech coding where the samples of speech signal are partitioned into frames. This indicates that the segmenting is not based on which quantization mode is selected.

At col. 3, lines 45-59, *Gersho* only discloses that "a highly efficient encoding of the excitation frame is achieved by directing processing to the windows themselves, and allocating all or nearly all of the available bits to code the regions inside the windows". *Gersho* only discloses which part of the excitation is used to encode, but not which quantization mode is selected. Furthermore, *Gersho* does not disclose or suggest that the partitioning of the samples into frames is based on which part of the excitation is encoded.

In Figure 5, *Gersho* depicts a ternary pulse coding AsB stage is used after the adaptive code book AbS stage. However, this has nothing to suggest that the partitioning of samples of speech signal into frames is based on whether a ternary pulse coding AsB stage is used.

At col.11, lines 4-16, *Gersho* discloses how the location of the window is positioned and that the location of the window is quantized in order to reduce the bit rate. However, this has nothing to suggest that the partitioning of samples of speech signal into frames is based on how the window is positioned and whether the window is quantized.

At col. 4, lines 33-39, *Gersho* discloses that the classes include voiced frames, unvoiced frames and transition frames and that the classes include strongly periodic frames, weakly periodic frames, erratic frames, and unvoiced frames. It is respectfully submitted that how frames are classified as voiced, unvoiced, weakly periodic frames or strongly periodic frames has nothing to do with selecting a quantization mode.

At col. 15, lines 35-36, *Gersho* discloses that if the Rate(m)=1, then the current frame is declared as a silent frame. Otherwise the current frame is declared as active speech. It is respectfully submitted that how a current frame is declared has nothing to do with selecting a quantization mode.

At col. 9, lines 63-65, *Gersho* discloses that a frame classifier sends two bits per basic frame to the speech decoder in the receiver to identify the classes. It is respectfully submitted that this has nothing to do with selecting a quantization mode and how the samples of speech signal are partitioned into frames.

As for claim 13, the Examiner states that *Gersho* discloses segmenting being carried out based on target accuracy (col. 9, lines 63-65 and col. 3, lines 45-49).

At col. 9, lines 63-65, *Gersho* discloses that a frame classifier sends two bits per basic frame to the speech decoder in the receiver to identify the classes. It is respectfully submitted that this has nothing to do with how the samples of speech signal is partitioned into frames.

At col. 3, lines 45-59, *Gersho* only discloses that "a highly efficient encoding of the excitation frame is achieved by directing processing to the windows themselves, and allocating all or nearly all of the available bits to code the regions insider the windows". *Gersho* only discloses which part of the excitation is used to encode. *Gersho* does not disclose or suggest that the partitioning of the samples into frames is based on target accuracy.

As for claim 14, the Examiner states that *Gersho* discloses that the segmenting step is carried out for providing a linear pitch representation in at least some of the segments (col.9,

lines 63-65; col.3, lines 45-49 and col.4, lines 50-62). It is respectfully submitted that, at col. 9, lines 63-65, *Gersho* discloses that a frame classifier sends two bits per basic frame to the speech decoder in the receiver to identify the classes. It is respectfully submitted that this has nothing to do with providing a linear pitch representation in some of the segments.

At col. 3, lines 45-59, *Gersho* only discloses that "a highly efficient encoding of the excitation frame is achieved by directing processing to the windows themselves, and allocating all or nearly all of the available bits to code the regions insider the windows". *Gersho* only discloses which part of the excitation is used to encoding. That has nothing to with providing a linear pitch representation in some of the segments.

At col.4, lines 50-55, *Gersho* discloses that the step of classifying uses a first classifying to classify a frame as being one of an unvoiced frame or a not-unvoiced frame, a second classifier for classifying a not-unvoiced frame as being one of a voiced frame or a transition frame. At col.4, lines 51-61, *Gersho* discloses that the frame are further partitioned into sub-frames, and at least one window within each sub-frame is positioned at a location that is a function of a pitch of the frame. This paragraph has nothing to do with providing a linear pitch representation in some of the segments.

As for claims 19 and 27, the Examiner states that *Gersho* discloses

an input for receiving audio data indicative of the parameters in the adjusted representation (input applied to element 14, Figure 3), and

a module responsive to the audio data for generating the audio signal based on the adjusted signals and the characteristics of the audio signal (Figure 3).

It is respectfully submitted that the input and the adjustment modules in claims 19 and 27 are components in a decoder for receiving audio data indicative of the parameters. In *Gersho*, Figure 3 shows an LP filter (14) in an encoder, and the input to the LP filter includes a speech signal and LP prediction parameters for tracking changes in the speech statistics. The speech and the update prediction parameters are not audio data indicative of the parameters as received in a decoder as claimed.

As for claims 20 and 28, the Examiner states that a person skilled in the art would record audio parameters in order to update the audio data for storage and retrieval. However, the

Examiner seems to refer to the updating of LP prediction parameters in the encoder as shown in Figure 3. This updating has nothing to do with the decoder as claimed in claims 20 and 28.

As for claims 21 and 29, the Examiner states that *Gersho* discloses that the audio data is transmitted through a communication channel and the input of the decoder is operatively connected to the communication channel for receiving the audio data (digital communications, col. 1, line 1 and Figure 3). It is respectfully submitted that lines 1-3 of col.1, are simply the title of the invention "Adaptive windows for Analysis-by-Synthesis CELP-type speech coding". Figure 3 shows an encoder. This is irrelevant to the features for a decoder as claimed in claims 21 and 29.

As for claim 22, the Examiner states that *Gersho* discloses a coding device as claimed. However, the Examiner fails to point out where *Gersho* discloses "said adjusting comprises the steps of segmenting the audio signal into a plurality of segments based on the characteristics of the audio signals". *Gersho* only discloses classifying the frames after the samples of speech signal are partitioned into frames.

As for claim 23, the Examiner states that *Gersho* discloses a quantization module, responsive to the adjusted representation, for coding the parameters in the adjusted representation (Figure 9). It is respectfully submitted that, in Figure 9, *Gersho* only shows that different encoders are used to encode the modified residual signal based on classification OCL(m).

As for claim 24, the Examiner states that *Gersho* discloses an output end, operatively connected to a storage medium, for providing data indicative of the coded parameters to the storage medium for storage (col.1, lines 64-65). At col. 1, line 65 to col. 2, line 1, *Gersho* discloses storing excitation candidates as vectors in a codebook and the coding method is referred to as code excited linear prediction (CELP). It is respectfully submitted that the excitation candidates are not audio data indicative of the coded parameters as provided by the encoder. The audio in claim 24 is speech data, whereas the excitation candidates stored in a codebook in the decoder are not speech data.

As for claim 31, the Examiner states that *Gersho* discloses a cell phone having a decoder as claimed. It is respectfully submitted that *Gersho* does not disclose or even suggest that said adjusting comprises the steps of segmenting the audio signal into a plurality of segments based on the characteristics of the audio signals. *Gersho* only discloses classifying the frames after the samples of speech signal are partitioned into frames.

As for claim 32, the Examiner states that *Gersho* discloses a decoder as claimed. However, the Examiner fails to show that *Gersho* discloses that the pitch contour data in the audio segment in time is approximated by a plurality of consecutive sub-segments in the audio segments, each of the sub-segments defined by a first end point and a second end point, and that the decoder has an input for receiving audio data indicative of the end points defining the sub-segments.

As for claim 34, the Examiner states that *Gersho* discloses that the audio signal contains sinusoidal components (col.3, lines 25-29) and the parameters include frequency values (Figure 1, element 68), amplitude values (col.3, lines 51-55) and phase values indicative of the sinusoidal components (Figure 1, element 76; col.3, lines 25-29).

At col.3, lines 25-29, *Gersho* discloses that the excitation signal within a sub-frame is constrained to be zero outside of selected intervals within the sub-frame, and these intervals are referred to as windows. Element 68 of Figure 1 is referred to a frequency synthesizer for providing the required frequencies to the receiver and transmitter. The frequency synthesizer 68 is a local oscillator. The frequencies provided by the frequency synthesizer is not the frequency values of the sinusoidal components of the speech signal.

At col.3, lines 51-55, *Gersho* discloses using three ternary valued amplitudes of 0, -1 and +1 to reduce the coding complexity. The ternary values are used in connection with the adaptive code book AbS (see Figure 5). This has nothing to with the amplitude values of the sinusoidal components of the speech signal from the microphone 72B in Figure 1 of *Gersho*.

Element 76 in Figure 1 of *Gersho* is an I/Q demodulator. I/Q demodulator is concerned with the modulated RF signal as received by the receiver 64. It has nothing to do with the phase value of the sinusoidal components of the speech signal from the microphone 72B in Figure 1 of Gersho.

As for claim 35, the Examiner states that *Gersho* discloses that the parameters include pitch (col.4, line 60), voicing, amplitude (col.3, lines 51-55) and energy of the audio signal (col.3, lines 42-44).

At col.4, line 56-64, *Gersho* discloses that the frames are further partitioned into sub-frames, and at least one window within each sub-frame is positioned at a location that is a function of a pitch of the frame. The location of a window has nothing to do with the pitch included in the parameters as claimed.

At col.3, lines 51-55, *Gersho* discloses using three ternary valued amplitudes of 0, -1 and +1 to reduce the coding complexity. The ternary values are used in connection with the adaptive code book AbS (see Figure 5). This has nothing to with the amplitude values of the speech signal as presented in the parameters.

At col.3, lines 42-44, *Gersho* discloses that classifying the speech signal includes a step of forming a smoothed energy contour from the residual signal and a step of considering a location of peaks in the smoothed energy contour in order to identify the location of the window (col.4, lines 64-67). *Gersho* does not disclose using energy as a parameter as claimed.

As for claim 36, the Examiner states that *Gersho* discloses that the parameters include pitch contour data (col.4, lines 60-61) containing a plurality of pitch values representative of an audio segment in time (col.4, lines 59-63, and col.2, lines 51-64).

At col.4, line 56-64, *Gersho* discloses that the frames are further partitioned into sub-frames, and at least one window within each sub-frame is positioned at a location that is a function of a pitch of the frame. The location of a window has nothing to do with the pitch contour data included in the parameters as claimed.

At col.2, lines 51-64, *Gersho* discloses that "it is a second object and advantage of this invention to provide a time-domain real-time speech coding/decoding system based on at least in part on a code excited linear prediction (CELP type algorithm, the speech coding/decoding system using adaptive windows". However, this second object has nothing to do with having parameters including pitch contour data containing a plurality of pitch values representative of an audio segment in time.

For the above reasons, claims 4, 5, 8-14, 20-25, 28, 29, 31, 32 and 34-36 are clearly distinguishable over the cited *Gersho* reference.

Furthermore, claims 3-14, 20, 21, 23-25, 28-30, 33-40 are dependent from claims 1, 19, 22, 26, 27 and 31 and recite features not recited in claims 1, 19, 22, 26, 27 and 31. For reasons regarding claims 1, 19, 22, 26, 27 and 31 above, it is respectfully submitted that claims 3-14, 20, 21, 23-25, 28-30, 33-40 are also distinguishable over *Gersho*.

At section 6, claims 15-18 are rejected under 35 U.S.C as being unpatentable over *Gersho* in view of *Gersho IEEE-96*.

It is respectfully submitted that claims 15-18 are dependent from claim 1 and recite features not recited in claim 1. For reasons regarding claim 1 above, claims 15-18 are also distinguishable over *Gersho* in view of *Gersho IEEE-96*.

As for new claims 41 – 48, they are dependent from claims 1, 19, 22, 26, 31 and 32 and recite features not recited in claims 1, 19, 22, 26, 31 and 32. For reasons regarding claims 1, 19, 22, 26, 31 and 32 above, it is respectfully submitted that claims 41 –48 are also distinguishable over the cited *Gersho* and *Gersho IEEE-96* references.

Furthermore, in the invention as claimed in claims 41 and 43, the audio signal comprises a plurality of frames and the audio signal in each frame has a waveform and wherein the further audio signal is produced in the decoding stage independently of the waveform. In contrast, *Gersho* discloses a CELP-type encoding method. As known by a person skilled in the art, a CELP coder is an example an Analysis-by-Synthesis (AbS) coder. A CELP coder performs waveform matching on the coder output using code excitation candidates and selecting the one minimizing given error criteria.

In the invention as claimed in claims 42, 44-48, not all the segments have equal segment length. Some segments contain fewer than 20 frames while some segments contains more than 20 frames (see Figure 3). In contrast, in *Gersho*, samples of speech signal are partitioned into frames.

Thus, Gersho does not disclose or even suggest the features as recited in claims 41-48.

CONCLUSION

Claims 1, 3 - 48 are allowable. Early allowance of all pending claims is earnestly solicited.

Respectfully submitted,

Kenneth Q. Lao
Attorney for the Applicant
Registration No. 40,061

WARE, FRESSOLA, VAN DER SLUYS
& ADOLPHSON LLP
Bradford Green, Building Five
755 Main Street, P.O. Box 224
Monroe, CT 06468
Telephone: (203) 261-1234
Facsimile: (203) 261-5676
USPTO Customer No. 004955